



## AN APPLICATION OF PEOPLE'S SENTIMENT FROM SOCIAL MEDIA TO SMART CITIES

Aplicación a las ciudades inteligentes del sentimiento de la gente en redes sociales



**Francisco-José Estévez-Ortiz, Antonio García-Jiménez and Peter Glösekötter**



**Francisco-José Estévez-Ortiz** is a doctor in Computer Science. His main research areas are wireless sensor networks and smart city infrastructure. From 2010 to 2012 he worked for *Cilab SL* (Granada, Spain) as a full-time researcher, after that, he moved to Germany. From 2012 to 2013 he worked for *Zodiac Aerospace* (Lübeck, Germany) as a full-time developer. In 2013, he moved to the *University of Applied Sciences of Münster* (Münster, Germany) and became a full-time researcher. Since then, his main research area is wireless sensor network as infrastructure for Smart Cities and low-powered embedded systems.

<http://orcid.org/0000-0002-7877-0115>

*Fachhochschule Münster, Raum: D-223  
Stegerwaldstrasse, 39. D-48565 Steinfurt, Germany  
fjestevez@ieee.org*



**Antonio García-Jiménez** is a full professor at the *Faculty of Sciences of Communication* in *King Juan Carlos University*. A large part of his research has been focused on knowledge organization and documentation and media. He has led and participated in different research projects related to systems of personalization, organization of web-information and the uses and risks of the Internet for minors. His latest works have been related to technological development and postgraduate communication studies in Spain.

<http://orcid.org/0000-0002-6860-446>

*Universidad Rey Juan Carlos  
Facultad de Ciencias de la Comunicación  
Camino del Molino, s/n. 28943 Fuenlabrada (Madrid), Spain  
antonio.garcia@urjc.es*



**Peter Glösekötter** received the diploma and Dr.-Ing. degrees from the *University of Dortmund*, Germany, in 1997 and 2002, respectively. From 2003 to 2006, he was associated with *Intel Research Lab*, Braunschweig, Germany. In 2006, he became full professor at the *Faculty of Electrical Engineering and Computer Science* at *Fachhochschule Münster*, Germany. As head of the *Semiconductor and Bus Lab*, his main research activities are embedded systems implementations of self-sustaining systems and architecture for emerging technologies.

<http://orcid.org/0000-0002-6937-468X>

*Fachhochschule Münster, Raum: D-223  
Stegerwaldstrasse, 39. D-48565 Steinfurt, Germany  
peter.gloesekoetter@fh-muenster.de*

### Abstract

This paper presents a method for analyzing social media and digital governance in smart cities. This research, after selecting a group of issues related to people's concerns about daily life in cities, makes use of *Afinn* lexicon as dictionary with sentiment score in order to recognize and measure the strength of individual words. The experiment retrieves different tweets classified by hashtags in a database and calculates an average sentiment. It involves some statistical processes for the recognition of the topics' sentiment: the calculation of the weighted average, the standard deviation, and the distribution among -5 to +5, based on the score range used by the lexicon. Finally, human evaluation is used to compare results from the automatic system.

## Keywords

Social media; Smart cities; Sentiment analysis; Public opinion; *Twitter*.

## Resumen

En este trabajo se presenta una propuesta basada en el “sentiment analysis” en medios sociales y orientada a la gobernanza digital en el marco de las ciudades inteligentes. Este sistema, después de seleccionar un grupo de temas relacionados con las preocupaciones de las personas acerca de la vida diaria en las ciudades, hace uso del léxico *Afinn* dado que el diccionario aplica una puntuación a cada palabra individual con el fin de reconocer y medir la fuerza de cada una. El experimento recupera tuits, clasificados por hashtags en una base de datos y, para cada uno de ellos, calcula el “sentimiento”. Esto implica el uso de procesos estadísticos para el reconocimiento del “sentimiento” como: el cálculo de la media ponderada, la desviación estándar y la distribución entre -5 a +5, el rango de puntuación usado por el léxico. Finalmente, se realiza una evaluación humana con el objetivo de comparar los resultados derivados del sistema automático.

## Palabras clave

Medios sociales; Ciudades inteligentes; *Sentiment analysis*; Opinión pública; *Twitter*.

Estévez-Ortiz, Francisco-José; García-Jiménez, Antonio; Glösekötter, Peter (2016). “An application of people’s sentiment from social media to smart cities”. *El profesional de la información*, v. 25, n. 6, pp. 851-858.

<https://doi.org/10.3145/epi.2016.nov.02>

## 1. Introduction

A contribution to smart cities and smart governance is presented in this paper in connection with social media by using large amounts of existing information. The underlying assumption is that social media has clear potential for the development of society, political institutions, and further organizations. For instance, the detection of mistakes and political irresponsibility is immediately reflected by citizens using social media.

This paper focuses on the application of sentiment analysis for opinion extraction. In general terms, sentiment analysis refers to the analysis (by means of natural language processing and computational linguistics techniques) focused on the opinions (positive, negative, or neutral), attitudes, and emotional tone in relation to an event, comment, person, object, or institution, applied in digital texts such as social media. Using sentiment analysis, this paper analyzes social media, such as social networking, microblogging, multimedia platforms, crowdsourcing platforms, collaborative tools, etc. The objective is to determine public opinion in terms of the local management of government. Although opinion mining and sentiment analysis are similar and belong to the same field of study, this paper is more concerned with research on attitudes, emotions, and perceptions rather than personal judgements. In addition, these techniques are employed to determine subjectivity in terms of positive or negative, social media monitoring, measurement of the strength of sentiment in texts, and working with linguistic tools in order to align this research to sentiment analysis to a greater extent.

The remainder of this work is structured as follows: an overview of the bibliography about smart cities, social media, sentiment analysis, and public opinion; analyzing the methodology of the presented proposal; describing, in detail, the experiment. Finally, the last section presents the conclusions.

## 2. State of art

This paper is focused on smart cities, sentiment analysis, and the relationship between public opinion and social media. These variables are analyzed and some related work is included for discussion.

### 2.1. Smart cities and social media

As pointed by **Neirotti et al.** (2014), the smart city aims to improve a citizen’s quality of life and has, therefore, increased in importance in both political and academic agendas. The expansion and growth of cities has resulted in traffic problems, pollution, and social inequalities (**Kim; Han**, 2012). At this point, technologies play a critical role.

A smart city provides a large number variables for analysis. For **Chourabi et al.** (2012), the key elements are: management and organization, technology, governance, political context, people and communities, the economy, infrastructure, and the environment. The connection to the Internet of Things (IoT) paradigm is evident. **Perera et al.** (2014) propose a sensing model as a service model applicable to smart cities with the aim of improving built-in cloud computing, participatory activity, and processes oriented to share and reuse information (reducing the cost of acquisition) to collect previously unavailable data, to increase the ability to access data from stakeholders, and to generate applications related to government actions in the context of smart cities.

Following up on studies from researchers such as **Criado, Sandoval-Almazán**, and **Gil-García** (2013), the impact that social media can have on innovation in government and in the public sector is observed. This fact is linked to transparency, citizen participation, and the increase in efficiency of administrative processes. Moreover, cost savings, good governance, and citizen satisfaction are involved. Overall, this influence is observed more in innovation in the field of interaction with citizens than in internal management.

Register for free at <https://www.scipedia.com> to download the version without the watermark

Thus far, social studies about smart cities have focused on topics such as opening city bureaucracy to the public, collaboration between public institutions, citizen participation in political decisions, maximizing the resources of operations, and information dissemination. In addition, **Oliveira and Welch (2013)** point to social media as technical and social systems where the interaction between technology and tasks is especially relevant; from this perspective, social media is similar to other computer-based communication tools, with a sufficient degree of flexibility to allow different uses. Empirical data drawn from their survey found that social media was used to disseminate information, allow or facilitate participation by citizens and external stakeholders, enable internal collaboration about work tasks, and obtain feedback on the quality of service. In addition, evidence suggests that social media is used for various purposes depending on the type of agency or government. In their paper (**Oliveira; Welch, 2013**) they showed the difficulty of taking into account all the variables that can be involved, such as the technology applied, the structures, the organizational process, and the texts analyzed.

**Riel, Popescu, and Guanlao (2014)** place social media into two categories: crises and normal situations. Regarding the former, social media can improve the dissemination of important information while also helping disaster management teams. For situations that are normal, social media enhances the interaction between citizens and governments. Examples include information requests by citizens (receiving responses in near real time), the diffusion of information about election activities, obtaining citizens' opinions about different subjects, and the effective involvement of citizens in caring for trash, trees, and fire hydrants. In the context of smart cities, citizens can also contribute to the decisions of public entities, generating knowledge on both sides.

The separation between crisis and normal situations poses implementation difficulties for sentiment analysis for this research. Nonetheless, this paper emphasizes the relevance of this group of techniques in order to detect citizens' opinions in real-time.

## 2.2. Sentiment analysis

This research is related to the categorization of opinions and expressions in social networks, linked to entities, facts, and properties. The idea is to use the expression of opinions and sentiments as a basic element for decision-making at different levels of administration. This method groups a set of computational techniques targeted at detecting opinions, emotions, feelings, beliefs, and/or evaluations (**Balahur; Mihalcea; Montoyo, 2014**). In connection with this paper other research has sought to monitor public opinion about specific issues (**Kim; Han, 2012; Stylios et al., 2010**), analyze candidates and social trends (**Ceron et al., 2014; Akcora et al., 2010**), and predict elections (**Tumasjan et al., 2011**).

It is useful to use the taxonomy of sentiment techniques oriented to dimension of private opinions proposed by **Bravo-Márquez, Mendoza and Poblete (2014)**:

1. Polarity: working with categorical variables whose possible values are positive, negative, and neutral. Lexical resources

used consist of lists of positive and negative words.

2. Strength: these methods and resources facilitate intensity levels according to a dimension of sentiment-polarity. They report from numeric scores that indicate the intensity or strength of positive and negative sentiments expressed in a text. Lexical resources configured by a list of words with intensity scores are used.

3. Emotion: these methods and resources are focused on the extraction of the emotion or mood of a text passage. The idea is to classify a message into an emotional category such as sadness, joy, and surprise. The key is in incorporating lists of words or expressions marked according to emotional states.

Also, **Roger and Kogan (2013)** emphasize the orientation of the word (positive, negative, or neutral) and the determination of both the subjectivity of the term and its strength. From their point of view, research about opinion mining is presented in three ways:

- a) development of linguistic resources for sentiment analysis as manually annotated corpus;
- b) implementation of algorithms for text analysis and classification according to semantic and subjective orientation, and;
- c) extracting text opinion, including different types of relationships with related content.

Their paper demonstrates the validity of the continuous monitoring of public opinion. However, it takes the same approach to various kinds of texts and messages.

**Kim and Kim (2014)** focus on monitoring public sentiment related to nuclear energy on *Twitter*. Their study is divided into four phases:

- a) cleaning the texts and removing the noise;
- b) cleaning and preprocessing of texts,
- c) construction of a dictionary of sentiments, and
- d) predicting the sentiment of tweets and scoring.

In the tracking phase, tweets that include "nuclear" terms or "nuclear power" in Korean are detected. In this case, the experiment provides inspiring results of this technique, also in terms of sentiment prediction and sentiment changes about a relevant social topic.

Also, **Villena-Roman, Luna-Cobos, and González-Cristóbal (2014)** present a system which stores, analyzes, and displays information from *Twitter* messages. In this case, the semantic tagging process is performed and takes into account that the texts are composed of short fragments that use abbreviations, emoticons, etc. First, the classification is carried out by a previously tested algorithm (based on rules filtering) and with two specific ontologies, in order to define the issues of the classification and other aspects such as location, events having places, disasters, and alerts. Secondly, it performs the extraction of entities (named entities, concepts, temporal expressions, and monetary terms) by a combination of natural language processing techniques. At the third level, a sentiment analysis in terms of positive, neutral, or negative is applied, considering each phrase

Register for free at <https://www.scipedia.com> to download the version without the watermark



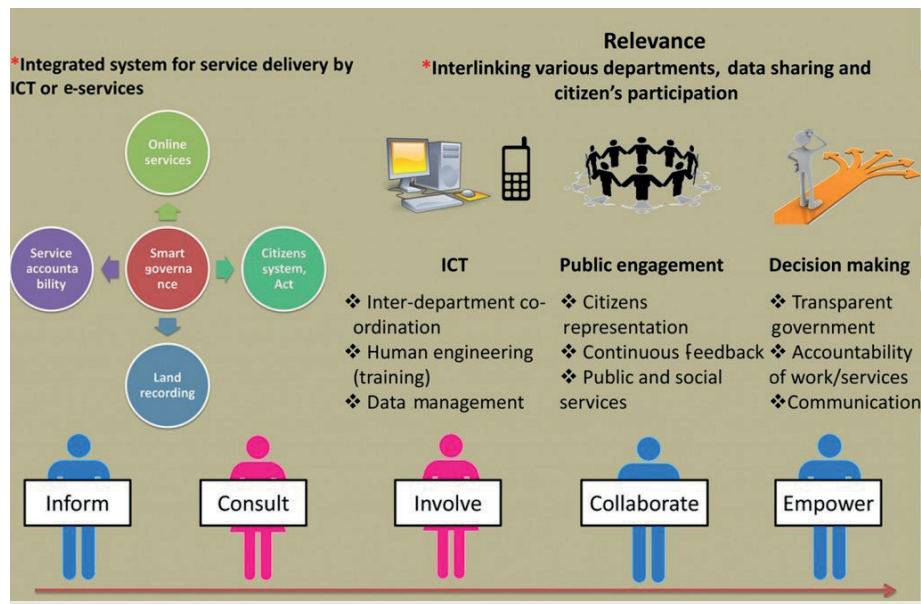


Figure 1. From the feelings and data to the decision-making. Image courtesy of the CEPT University, Gujarat, India

and the combination of phrases. It also takes into account the demographic characteristics: type of user, user's gender, and the user's age range. This research presents a relevant global architecture in order to apply sentiment analysis.

In the proposal of Ceron *et al.* (2014) the HK method, which includes two phases, is applied. The first one consists of reading and coding a subsample of documents downloaded from the Internet by human coders. This subsample is converted into a training set that will be used by the HK algorithm to classify documents. In the second stage, statistical analysis is automated by the HK algorithm and extends all the documents analyzed, capturing suitable opinions expressed on the Web. In this case the correlation between the analysis of social media preferences and the results of traditional media are presented. The prediction capability for both "single-issue" elections and for more complex elections is shown.

Another aspect that has been useful in this context has to do with sentiment analysis focused on news. For example, in the research of Balahur *et al.* (2010), the idea that opinion mining is different from other types of mining is highlighted. One of the most outstanding features is the delimitation of labeled opinion analysis that is expressed explicitly, especially in texts that do not require an in-depth interpretation.

### 2.3. Social media and public opinion

According to Anstead and O'Loughlin (2015), the role of social media is undeniable in situations of political confrontation or election debates. The monitoring of social media allows the analysis of social dynamics through public opinion, particularly during elections.

However, this method of testing public opinion raises some

doubts. In fact, the report of the American Association for Public Opinion Research (Murphy *et al.*, 2014) about this issue explains that the use of social media as a source of information for research requires validation through other forms of information sources. The key is to avoid any errors or poor quality of information which, in both cases, could result in misinterpretations.

Another question is whether social media can be representative of the global sentiment of the population, or, to be more specific, of users of the Internet and social media. In the context of social media research, the problem is the possibility of creating a "frame of social media users" from which a sample of

individuals can be used for research.

It should also be noted that not everyone has access to the Internet. And access to the Internet and social media can vary due to many circumstances: the different social media platforms, the type of activities on social media sites, the time and place of access, or even demographic and personal characteristics. Thus, Xiong and Liu (2014) note that public opinion often evolves into a state dominated by a single opinion, but without reaching consensus. And they also indicate that the agents involved in these processes are reluctant to change their minds.

In this regard, Murphy *et al.* (2014) suggest to combine this type of research with surveys and statistics, which provide an alternative perspective, either quantitative or narrative, to traditional analysis. Ethical implications, which result from the monitoring of social media like impartiality, transparency, accuracy, and quality of the used measurements, should be also solved.

### 3. Methodology

This work presents the interaction between the sentiments of society, internet opinions, and local governance. This kind of system allows a fast response against heavy problems or critical scenarios. Figure 2 shows the implemented algorithm for that purpose. This section details the design process.

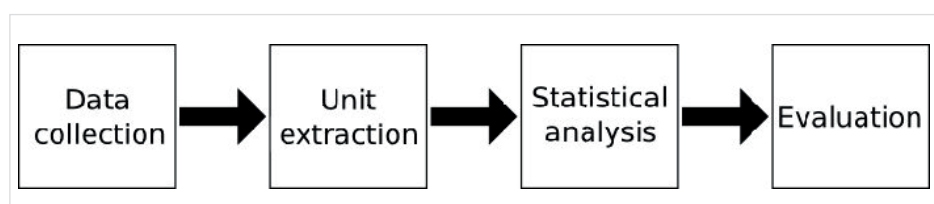


Figure 2. Phases of the proposed system

enviroment, pollution, roadtraffic,  
 security, crime, delinquency, arrest,  
 police, infrastructure, politics,  
 election, electoralresult, candidate,  
 militant, citycouncil, tourism,  
 emergency, health, socialservice,  
 townplan, developmentplan, economy,  
 publicfinance, tax, budget, transport,  
 publictransport, culture, organization,  
 society, education

Figure 3. List of words used as topics

### 3.1. Data collection and unit extraction

One of the aims of this study is the analysis of people's sentiment. A way to recognize those sentiments is by using a data collection, mainly focused on written expression. Written expressions use rich language, allowing for linguistic unit recognition. Most of the expressions can be acquired from *Facebook* or *Google+*'s status, comments, tweets, posts on online media, and blogs. The data collection can have several inputs, but it should be consistent, clearly written and organized by topic.

The topics are manually selected and only English is used. The future objective is automatic topic selection and multilingualism. The idea behind topic selection is to collect issues related to people's concerns about daily life in cities, taking into account news from various media sources. The use of topics allows the adaptability of this work to other areas. At the moment, this study is focused on smart cities, social media, and digital governance, thus the words are all subject-related with those topics. Those words used as topics are listed in Figure 3.

The recognition of linguistic units should be developed by a tool that recognizes and measures the strength of individual words. This proposal makes use of the *Afinn* lexicon as a dictionary with sentiment score. This lexicon is chosen because it is centered in the typical language of social media with obscene words, acronyms, and web jargon (Nielsen, 2011). Moreover, *Afinn* uses a score or measurement range that is very suitable for the purpose of this paper. This score catalogs positive words from 1 to 5, and negative words from -1 to -5. *Afinn* is a well known resource between researchers and for this paper, a good starting point with language recognition.

### 3.2. Statistical analysis

The system for determining the strength of written expressions is based on individual statistical processing, in order to obtain enough information about written expressions and topics.

Once that the written expressions are classified by topic, a general dataset is compiled and the different opinions are recognized (as individual units in the dataset), as explained in section 3.1. These opinions are separated into individual

words, where each unit is represented in a vector with the used terms (also known as *word profile*). After that, an average is calculated using the score from the *Afinn* lexicon (-5 to +5), counting the matching individual words and making a score average (known as *tweet average*). From the distribution function it is trivial to derive standard deviation and the weighted average. The weighted average is based on equation (1), where  $n$  is the number of matching words,  $i$  is the *Afinn*'s score,  $m$  is the number of possible intervals for the distribution and  $W$  the weight of that interval in the distribution.

$$\bar{X}_T = \frac{\sum_{i=1}^n \left( \frac{\sum_{i=1}^n V_i}{n} \times W \right)}{m} \quad (1)$$

The ratio is calculated based on a confidence interval of a 95%. The tweets of a topic are analyzed and, if they are into the confidence interval, they are stored. Otherwise they are not considered. Finally, the polarity of a topic is calculated analyzing the sentiment distribution of the tweets for the sample in the confidence interval.

## 4. Experiment and results

*Twitter* is the social media platform selected for analysis. *Twitter* is a microblogging service that allows people to share opinions and information related to any topic or event. The textual information was aggregated and analyzed in order to determine states of public opinion. It has been suggested that *Twitter* is primarily used for information dissemination to link like-minded people rather than to discuss any issue. Users often relay information instead of dealing with an idea. Users of *Twitter* have an opinion on an issue before publication on the matter and it appears unlikely that they change their mind in tweets (Smith et al., 2013).

The method for analysis for this study was developed using a self-developed java program and the library *Twitter4J*, which is an unofficial library for the *Twitter* API. Basically, the program analyzes a database of hashtags and sends request about those hashtags to the *Twitter* API, using *Twitter4J*. This library is a well known tool within the developer's community and it has been previously used by various research groups.

The keywords used as topics for the search (hashtags in *Twitter*) were retrieved from more than 200,000 tweets, in order to represent media activity. Other techniques such as topic modelling will be evaluated in next proposals. Those tweets were classified by hashtags in a database and, for everyone, an average sentiment was calculated.

Specifically, this experiment involved some statistical processes for the recognition of a topic's sentiment. First, a statistical analysis of the topic was conducted. This statistical analysis involved the calculation of the weighted average, the standard deviation, and the distribution among -5 to +5 (the score range for the *Afinn* lexicon). Second, an estimation was calculated of the highest and lowest limit for a confidence interval, the ratio of tweets in this confidence interval, and the polarity of the topic. Later, a console-based interface showed the basic statistical information and the

Register for free at <https://www.scipedia.com> to download the version without the watermark

topic was respectively classified as positive, neutral, or negative, and showed more detailed information about the negative ones.

The results varied depending on the topics. The software made for this work stored the statistical data of the topics in a file, which allowed for later review of the results.

The results of this work were shown in a text-based manner through a java-based multi-platform terminal software. In the following section some examples are classified by topic, in order to give a reliable representation of the tweet sample:

### Environment

- Birds are indicators of the environment. If they are in trouble, we know we'll soon be in trouble. Roger Tory Peterson #environment
- How #polluted could be #NYC in a 99% #humidity day? #contamination #environment #agressivity #aggressiveEnvironment
- Then we have those who say, "fuck the #environment" see, @GOP to promote consumerism & the exploitation of resources beyond sustainability.

### Pollution

- I just wake up. I went outside and look at what o found. #pollution!? #beauty? #newyork #nyc #morning <https://instagram.com/p/-oLi4QCORB/>
- And this is WHY you should not swim in Hudson River! #pollution #hudsonriver #nyc #manhattan! <https://instagram.com/p/653Oj9xK7n/>
- Testifying now #NYC City Hall on the #ecological impact of #light #pollution @IDADarkSky @SaveBillionBird

### Traffic

- Accident in #Edgewater on Edgewater NB near Gorge Rd, stop and go traffic back to Bulls Ferry Rd, delay of 1 min #traffic
- Green light for carriages. #carriage #horse #carriageride
- On a #gridlock alert day in #NYC working on an hour to go 17 blocks. #Traffic #nbc4ny

### Economy

- In revitalizing upstate #NY's #economy, @NYGovCuomo has also been quietly implementing a #green revolution. <http://huff.to/1NFg3J4>
- No wage growth, no interest hike. McJobs != real economic growth. #Fed #economy
- Seriously? #nyc #parkingmeters #cars #money #economy #saturday @ Queens, NY <http://instagram.com/p/yiasSEMIL8/>

Using tweets like these, the software identified every single word as a token, and compared each one with the elements of *Afinn* lexicon. If a word match was found, the value of that word was saved. The matchings among the tweet's tokens were summed up. When that processing was finished, the tweet's average was calculated. The process continued until all the tweets of a specific topic were analyzed. After that, the distribution was calculated. From the distribution function, it was trivial to derive standard deviation

#delinquency Average: 0.2 Standard deviation: 0.303 Polarity: Negative Ratio: 90%	#roadtraffic Average: 0.3 Standard deviation: 0.746 Polarity: Negative Ratio: 86%
#publicfinance Average: -0.24 Standard deviation: 0.799 Polarity: Neutral Ratio: 76%	#townplanning Average: 0.31 Standard deviation: 1.095 Polarity: Neutral Ratio: 19%

Figure 3. Example of results from the 200,000-tweets sample

and the weighted average. The statistical processing continued calculating the confidence interval, for a 95%-confidence. This confidence interval calculated the highest and lowest limit. Afterwards the ratio was calculated, based on the number of tweets whose sentiment average was contained in the confidence interval. Finally, the polarity was calculated analyzing the sentiment distribution of the tweets that were contained in the confidence interval. Figure 3 showed an example of the results for the topics covered in the 200,000-tweets sample.

For comparison purpose, a small sample was formed from 40 tweets of the original 200,000-tweets sample, which were related to the topics *environment*, *pollution*, *traffic*, and *economy*. The sample was specifically focused on New York City, and tweets were only from that area. The tweet selection was manually developed, choosing 10 tweets for each topic and checking the provenance of each one. Trying to work with an actual sample, all the results were limited to January 2015.

The statistical results for each topic were:

### Environment

Average: 0.03  
Standard deviation: 0.504  
Polarity: Negative  
Ratio: 60%

### Pollution

Average: 0.03  
Standard deviation: 1.223  
Polarity: Negative  
Ratio: 70%

### Traffic

Average: -0.40  
Standard deviation: 0.489  
Polarity: Neutral  
Ratio: 70%

### Economy

Average: -0.50  
Standard deviation: 0.789  
Polarity: Neutral  
Ratio: 70%

Obviously there were tweets which weren't related to the topic, but where the topic word was used. It was difficult to automatically clean up those tweets from the collection because they referred, somehow, to that topic.

Register for free at <https://www.scipedia.com> to download the version without the watermark



In order to verify the results obtained by the automatic system, an online survey was developed using exactly the same 40-tweets sample used by the automatic system. This time, humans classified the average tweet sentiment, in the range of -5 to +5. Figure 4 shows an extract from the survey. This survey is still available under

<https://es.surveymonkey.com/r/8GZMFPD>

The tweets are randomly shown, the data about user ID was not collected (it was filled in as an anonymous user), and the survey is public.

Based on the results of 11 people, who filled out the survey, an average between the results was calculated and was used as an average human opinion for each tweet. Afterward, the same process was used and calculated the following parameters. The results are classified by topic:

#### Environment

Average: 0.545

Standard deviation: 1.755

Polarity: Neutral

Ratio: 60%

#### Pollution

Average: -1.227

Standard deviation: 1.369

Polarity: Negative

Ratio: 70%

#### Traffic

Average: -0.354

Standard deviation: 0.989

Polarity: Neutral

Ratio: 70%

#### Economy

Average: -0.309

Standard deviation: 1.716

Polarity: Neutral

Ratio: 70%

When comparing the results from the automatic system and human analysis, it appeared to be similar in terms of global results. In addition, a pertinent ratio (0.7) of intercoder reliability measure was observed. There was only a topic, #environment, where it was possible to find differences. These differences were probably a result of the different analysis that the automatic system made. In the automatic system, the impact of the keywords (understanding keyword as the word that changes the meaning of a tweet) was distributed among the rest of the tweet, whereas for the human classification, those keywords were the most relevant factor in the tweet, making no distribution among the rest. That was also the reason particular variables were given different results. Anyway, this work demonstrates the capabilities of automatic systems for sentiment analysis, besides its suitability in the context of smart cities.

## 5. Conclusion/discussion

The results of these experiments show the promising application of sentiment analysis to smart cities and digital governance. In this way, it is possible to define public opinion

in terms of local management, with linguistic and statistical improvements.

In addition, it is a useful and necessary a feedback channel, where the system itself can dynamically modify the actual policies or give recommendations. Through this system, the government, in particular local governments, can use it as feedback tool for their policies, allowing government to know the people's opinion about policies or their application. Please note, this research does not analyze whether social media can be representative of global sentiment and has been applied, thus far, to a local context.

Sentiment analysis was used in this study to present a valid technique that can enable smart cities to detect opinion/sentiment in real time, based on perceptions of citizens about one of the analyzed topics. For instance, the increase in negative tweets in relation to traffic or pollution allows city government implement a concrete measure in almost real time.

Second, this proposal is a useful tool to perceive the state of opinion of citizens on a global basis and about concrete events, and their correspondent changes, without connection with nuances.

Third, in our opinion, it requires a greater degree of refinement and mixing with other techniques to improve operative results in terms of different kinds or opinion and methods to categorize tweets.

Finally, there are several ways to further develop this research area, as an example, test another lexicon and try to obtain better opinion recognition, or develop a subjectivity measurement, and try to reduce the impact of the underlying negativity from the social media. These possible avenues of research would give a better idea of the vast possibilities of this research, conducted here and its reach.

## Acknowledgment

The authors would like to thank Tor Lund-Larsen for the fruitful discussions and contributions to this work.

## 6. References

- Akcora, Cuney-Gurca; Bayir, Murat-Ali; Demirbas, Murat; Ferhatosmanoglu, Hakan (2010). "Identifying breakpoints in public opinion". In: *1st Workshop on social media analysis*, pp. 62-66.  
[http://snap.stanford.edu/soma2010/papers/soma2010\\_9.pdf](http://snap.stanford.edu/soma2010/papers/soma2010_9.pdf)
- Anstead, Nick; O'Loughlin, Ben (2015). "Social media analysis and public opinion. The 2010 UK general election". *Journal of computer-mediated communication*, v. 20, n. 2, pp. 204-220.  
<http://eprints.lse.ac.uk/60796>  
<https://doi.org/10.1111/jcc4.12102>
- Balahur, Alexandra; Mihalcea, Rada; Montoyo, Andrés (2014). "Computational approaches to subjectivity and sentiment analysis: presented and envisaged methods and applications". *Computer speech and language*, v. 28, n. 1, pp. 1-6.  
<https://doi.org/10.1016/j.csl.2013.09.003>

Register for free at <https://www.scipedia.com> to download the version without the watermark

**Balahur, Alexandra; Steinberger, Ralf; Kabadjov, Mijail; Zavarella, Vanni; Van-der-Goot, Erik; Halkia, Martina; Pouliquen, Bruno; Belyaeva, Jenyua** (2010). "Sentiment analysis in the news". In: *Procs of the 7<sup>th</sup> conf on intl language resources and evaluation*, pp. 2216-2220.  
<https://arxiv.org/abs/1309.6202>

**Bravo-Márquez, Felipe; Mendoza, Marcelo; Poblete, Bárbara** (2014). "Meta-level sentiment models for big social data analysis". *Knowledge-based systems*, v. 69, pp. 86-99.  
<https://goo.gl/C4c7yT>  
<https://doi.org/10.1016/j.knosys.2014.05.016>

**Ceron, Andrea; Curini, Luigi; Iacus, Stefano M.; Porro, Giuseppe** (2014). "Every tweet counts? How sentiment analysis of social media can improve our knowledge of citizens' political preferences with an application to Italy and France". *New media & society*, v. 16, n. 2, pp. 340-358.  
<https://goo.gl/olqoME>  
<https://doi.org/10.1177/1461444813480466>

**Chourabi, Hamed; Nam, Taewoo; Shawn, Walker; Gil-García, J. Ramón; Mellouli, Sehl; Nahon, Karine; Pardo, Teresa A.; Scholl, Hans-Jochen** (2012). "Understanding smart cities: an integrative framework". In: *45<sup>th</sup> Hawaii intl conf on system sciences*. IEEE, pp. 2289-2297.  
[https://www.ctg.albany.edu/publications/journals/hicss\\_2012\\_smartcities/hicss\\_2012\\_smartcities.pdf](https://www.ctg.albany.edu/publications/journals/hicss_2012_smartcities/hicss_2012_smartcities.pdf)

**Criado, J. Ignacio; Sandoval-Almazán, Rodrigo; Gil-García, J. Ramón** (2013). "Government innovation through social media". *Government information quarterly*, v. 30, n. 4, pp. 319-326.  
[https://www.researchgate.net/publication/259131981\\_Government\\_Innovation\\_Through\\_Social\\_Media](https://www.researchgate.net/publication/259131981_Government_Innovation_Through_Social_Media)  
<https://doi.org/10.1016/j.giq.2013.10.003>

**Kim, Hyung-Min; Han, Sun-Sheng** (2012). "City profile: Seoul". *Cities*, v. 29, n. 2, pp. 142-154.  
[https://www.academia.edu/9584899/City\\_profile\\_Seoul](https://www.academia.edu/9584899/City_profile_Seoul)  
<https://doi.org/10.1016/j.cities.2011.02.003>

**Murphy, Joe; Link, Michael W.; Childs, Jennifer H.; Tesfaye, Casey L.; Dean, Elizabeth; Stern, Michael; Pasek, Josh; Cohen, Jon; Callegaro, Mario; Harwood, Paul** (2014). *Social media in public opinion research. Report of the Aaport task force on emerging technologies in public opinion research*. American Association for Public Opinion Research.  
[https://www.aapor.org/AAPOR\\_Main/media/MainSiteFiles/AAPOR\\_Social\\_Media\\_Report\\_FNL.pdf](https://www.aapor.org/AAPOR_Main/media/MainSiteFiles/AAPOR_Social_Media_Report_FNL.pdf)

**Neirotti, Paolo; De-Marco, Alberto; Cagliano, Anna-Corinna; Mangano, Giulio; Scorrano, Francesco** (2014). "Current trends in smart city initiatives: Some stylised facts". *Cities*, v. 38, pp. 25-36.  
<https://doi.org/10.1016/j.cities.2013.12.010>

**Nielsen, Finn-Årup** (2011). "A new ANEW: Evaluation of a

word list for sentiment analysis in microblogs". In: *Procs of the ESWC2011 Workshop on 'Making sense of microposts': Big things come in small packages*. CEUR Workshop Proceedings, pp. 93-98. Heraklion, Crete, Greece, May 30.  
<https://arxiv.org/abs/1103.2903>

**Oliveira, Gustavo-Henrique-Maultasch; Welch, Eric W.** (2013). "Social media use in local government: Linkage of technology, task, and organizational context". *Government information quarterly*, v. 30, n. 4, pp. 397-405.  
<https://goo.gl/A4JMK>  
<https://doi.org/10.1016/j.giq.2013.05.019>

**Perera, Charith; Zaslavsky, Arkady; Christen, Peter; Georgakopoulos, Dimitrios** (2014). "Sensing as a service model for smart cities supported by internet of things". *Transactions on emerging telecommunications technologies*, v. 25, n. 1, pp. 81-93.  
<https://arxiv.org/abs/1307.8198>  
<https://doi.org/10.1002/ett.2704>

**Riel, Arthur J.; Popescu, Denisa; Guanlao, Luisita** (2014). "Social data mining and knowledge flows between government and its citizenry in crisis and normal situations". In: *WIMS'14, Procs of the 4<sup>th</sup> intl conf on web intelligence, mining and semantics*.  
<https://doi.org/10.1145/2611040.2611090>

**Roger, Sandra; Kogan, Pablo** (2013). "Agentes inteligentes en ambientes dinámicos: Análisis de opinión". In: *XV Workshop de investigadores en ciencias de la computación*, Paraná, Entre Ríos, pp. 860-864.  
<http://sedici.unlp.edu.ar/handle/10915/27321>

**Smith, Laura M.; Zhu, Linhong; Lerman, Kristina; Zozareva, Zornitsa** (2013). "The role of social media in the discussion of controversial topics". In: *SocialCom'13. Procs of the 2013 ACM workshop on SocialCom*, pp. 215-223.  
<http://www.isi.edu/integration/people/lerman/papers/Smith13socialcom.pdf>  
<https://doi.org/10.1109/SocialCom.2013.41>

**Stylios, George; Christodoulakis, Dimitris; Besharat, Jerjes; Vonitsanou, Maria-Alexandra; Kotrotsos, Ioanis; Koumpouri, Athanasia; Stamou, Sofia** (2010). "Public opinion mining for governmental decisions". *Electronic journal of e-government*, v. 8, n. 2, pp. 203-214.  
<http://www.ejeg.com/issue/download.html?idArticle=209>

**Tumasjan, Andranik; Sprenger, Tim O.; Sandner, Philipp G.; Welp, Isabell** (2011). "Election forecasts with Twitter: How 140 characters reflect the political landscape". *Social science computer review*, v. 29, n. 4, pp. 402-418.  
<https://doi.org/10.2139/ssrn.1833192>

**Villena-Román, Julio; Luna-Cobos, Adrián; González-Cristóbal, José-Carlos** (2014). "Análisis semántico de la opinión de los ciudadanos en redes sociales en la ciudad del futuro". *Procesamiento del lenguaje natural*, n. 53, pp. 159-162.  
[http://www.taln.upf.edu/pages/sepln2014/projects/edited\\_paper\\_32.pdf](http://www.taln.upf.edu/pages/sepln2014/projects/edited_paper_32.pdf)

**Xiong, Fei; Liu, Yun** (2014). "Opinion formation on social media: An empirical approach". *Chaos*, n. 24.  
<https://doi.org/10.1063/1.4866011>

Register for free at <https://www.scipedia.com> to download the version without the watermark